PRIFYSGOL BANGOR / BANGOR UNIVERSITY

In silico analysis of the fragrance gene (badh2) in Asian rice (Oryza sativa L.) germplasm and validation of allele specific markers

Withana, Withanage Vidyani Erandika ; Kularathna, Rathanyaka Maudiyanselage Ramesha Eshani ; Kottearachchi, Nisha Sualri ; Kekulandra, Deepthika S. ; Weerasena, Jagath; Steele, Katherine

Plant Genetic Resources

DOI:
10.1017/S1479262120000015

Published: 01/04/2020

Peer reviewed version

Cyswllt i'r cyhoeddiad / Link to publication

Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):
Withana, W. V. E., Kularathna, R. M. R. E., Kottearachchi, N. S., Kekulandra, D. S., Weerasena, J., & Steele, K. (2020). In silico analysis of the fragrance gene (badh2) in Asian rice (Oryza sativa L.) germplasm and validation of allele specific markers. *Plant Genetic Resources*, *18*(2), 71-80. https://doi.org/10.1017/S1479262120000015

Hawliau Cyffredinol / General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal ?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

04. May. 2023

In silico analysis of the fragrance gene (_Badh2_) in Asian rice (_Oryza sativa_ L.) germplasm and validation of allele specific markers

Running title: **_In silico_ analysis of fragrant gene in Asian rice**

Authors:

Name: Withanage Vidyani Erandika Withana

Address: Department of Biotechnology, Faculty of Agriculture and Plantation Management, Wayamba University of Sri Lanka, Makandura, Gonawila (NWP), 60170, Sri Lanka.

Name: Rathanyaka Maudiyanselage Ramesha Eshani Kularathna

Address: Department of Biotechnology, Faculty of Agriculture and Plantation Management, Wayamba University of Sri Lanka, Makandura, Gonawila (NWP), 60170, Sri Lanka.

Name: Nisha Sualri Kottearachchi

Address: Department of Biotechnology, Faculty of Agriculture and Plantation Management, Wayamba University of Sri Lanka, Makandura, Gonawila (NWP), 60170, Sri Lanka.

Name: Deepthika S. Kekulandara

Address: Rice Research and Development Institute, Batalagoda, Sri Lanka.

Name: Jagath Weerasena

Address: Institute of Biochemistry, Molecular Biology and Biotechnology, University of Colombo, Kumaratunga Munidasa Mawatha, Colombo 03, Sri Lanka.

Name: Katherine A. Steele

Address: School of Natural Sciences, Bangor University, Bangor, Gwynedd LL57 2UW, United Kingdom.

Corresponding Author: Nisha Sualri Kottearachchi
Email: kottearachchins@yahoo.com, ORCID ID: 0000-0002-7406-6948

***In silico* analysis of the fragrance gene (*Badh2*) in Asian rice (*Oryza sativa* L.) germplasm and validation of allele specific markers**

Abstract

Badh2 of rice is considered to be the major gene responsible for the fragrance in rice. The wild type *badh2* allele encodes betaine aldehyde dehydrogenase 2 (BADH2) enzyme while the mutated version of *badh2* gene encodes non-functional BADH2 enzyme that leads to the accumulation of 2-acetyle-1-pyrroline (2AP), the principal fragrant compound in rice. There are many mutated recessive alleles causing fragrance in global rice germplasm, although the *badh2.1* allele present in *Basmati* type rice is the most well-known among breeders. In this study, we attempted to reveal potential fragrance causing mutations, and the respective varieties carrying them, through *in silico* analysis based on the sequences available in the Rice SNP-Seek-Database of International Rice Research Institute. The sequences of 1878 rice accessions from 22 countries were analyzed to identify mutations in each exon of *badh2* comparatively with the non-fragrant 'wildtype' GenBank sequence in Nanjing11, *Oryza sativa indica* (EU770319.1). Results revealed that 63 varieties from 12 countries possessed the most prevalent allele, *badh2.1* having an 8 bp deletion and 3 SNPs in the 7th exon. The second most prevalent allele in genotypes from Asia was *badh2.7* having a 'G' insertion in the 14th exon. A novel allele with a T deletion in 9th exon was detected in a Thai rice accession. Rice varieties containing either *badh2.1* or *badh2.7* alleles could be identified with DNA markers for *badh2.1* (*frg*) and *badh2.7* (*Bad2.7CAPS*). The marker, *Bad2.7CAPS*, co-segregated with the fragrance phenotype in two crosses, confirming the possibility of employing it in marker assisted breeding.

***Keywords*:** Aromatic rice, Bioinformatics, Marker assisted selection, Plant breeding,

Introduction

Aroma in rice is an important trait that determines the market price of rice (Giraud *et al.*, 2014). There are hundreds of volatile compounds causing fragrance of rice. Among them 2-acetyl-1-pyrroline (2AP) is regarded as the major aromatic compound which contributes for the fragrance in *Basmati* type rice (Sood and Siddiq, 1978; Buttery *et al*., 1982; Buttery *et al*., 1983). From linkage mapping studies Pinson *et al*. (1994); Lorieux *et al*. (1996); Cordeiro *et al.* (2002) and Bradbury *et al*. (2005a) determined that the major gene responsible for fragrance through elevated levels of 2AP, is located on chromosome 8. Bradbury *et al*. (2005b) reported that presence of an eight base pair deletion and three SNPs in exon 7 of the betaine aldehyde dehydrogenase 2 (*badh2*) gene on Chromosome 8 produces non-functional variant of the BADH2 enzyme leading to accumulation of 2AP in *Basmati* type rice. The particular allele containing the eight base pair deletion and three SNPs in exon 7 was named as *badh2.1* by Kovach *et al*. (2009) and as *badh2-E7* by Bindusree *et al*. (2017).

The complete *badh2* gene comprising of 15 exons and 14 introns, encodes 503 amino acids (Kovach *et al*., 2009). Kovach *et al*. (2009) have reported that ten possible allele types (*badh2.1* to *badh2.10* containing different mutations in exons) that contribute to fragrance in rice (Table 1). In addition, the presence of another several mutations have been reported (Shao *et al*., 2011; Shao *et al*., 2013 and He *et al*., 2015) in fragrant rice varieties from different geographical areas (Table 1).

Chen *et al.* (2008) and Sakthivel *et al.* (2009) have predicted that biosynthesis of 2AP occurs via a polyamine pathway with the involvement of the wildtype BADH2 enzyme. If the BADH2 enzyme is functional gamma-aminobutyric acid (GABA) is produced giving phenotypes without fragrance, conversely if the BADH2 enzyme is non-functional, the pathway diverts to the production of 2AP. A supporting evidence reported by Hinge *et al*. (2016) demonstrated that reduced expression of the *badh*2 gene dose, increases 2AP accumulation in rice grains. The *badh2.1* allele encodes a truncated non-functional BADH2 protein with 251 amino acids due to a premature stop codon.

Bradbury *et al*. (2005b) developed a marker known as *frg* comprising of 4 primers (*ESP*, *IFAP*, *INSP* and *EAP*) to differentiate the *badh2.1* using multiplex PCR. Dissanayaka *et al*. (2014) developed a Cleaved Amplified Polymorphic Sequence (CAPS) marker known as *Bad2.7CAPS*, to differentiate the *badh2.7* allele that possesses "G" insertion in exon 14. There is a need for employing user friendly DNA markers to detect different functional mutations of *badh2* gene to accelerate rice breeding. In this study, we attempted to detect potential fragrance causing mutations, and the respective varieties carrying them, through *in-silico* analysis using *badh2* sequences available at Rice SNP-Seek-Database of International Rice Research Institute (Alexandrov *et al*., 2014). As only few breeding efforts have been reported on utilization of *badh2.7* allele, we validated the *Bad2.7CAPS* marker as a functional marker using several traditional varieties and breeding lines.

Materials and methods

Extraction of badh2 sequences from Rice SNP-Seek Database

Sequences of *badh2* in rice from Asian countries were obtained from Rice SNP-Seek Database developed by IRRI (http:// snp-seek.irri.org) using the locus number Os08g0424500 located on chromosome 8 (20379823 bp - 20385975 bp). A total of 1878 rice accessions representing 22 countries of Asia were analyzed: one from Afghanistan, 170 from Bangladesh, 19 from Butan, 51 from Cambodia, 253 from China, 378 from India, 226 from Indonesia, two from Iran, 34 from Japan, 120 from Laos, 64 from Malaysia, 67 from Myanmar, 42 from Nepal, one from North Korea, 32 from Pakistan, 146 from Philippines, 29 from South Korea, 47 from Sri Lanka, 26 from Taiwan, 134 from Thailand, one from Tibet and 35 from Vietnam.

Mutation identification, multiple sequence alignment and amino acid sequence analysis

All previously reported mutations in *badh2* exon sequences (Table 1 and Figure 1) leading to non-functional BADH2 enzyme were considered in the analysis. Using MEGA 7 sequence analysis software (Kumar *et al*., 2016) the sequence of each rice accession was aligned with the complete cDNA of *Oryza sativa indica* non-fragrant variety Nanjing11 (www.ncbi.nlm.nih.gov/nuccore/eu770319.1). For each accession with a variant mutant sequence the whole predicted coding sequence of *badh2* was translated into amino acid sequence using the ExPASy-Translate tool (https://web.expasy.org/translate/) to check any predicted functional change in amino acid sequence compared to the wild type amino acid sequence translated from GenBank Acc. No. EU770319.1.

Plant materials used for screening of fragrant alleles

Seven aromatic varieties, two non-aromatic varieties and nineteen advanced breeding lines derived from two rice crosses, *Iginimitiya* × *Bg 300* and *Suwandel* × *Bg 360* were obtained from Rice

Research and Development Institute (RRDI), Batalagoda, Sri Lanka (Table 3). Phenotyping for fragrance in the advanced lines had been done previously by a sensory evaluation method as reported by Sood and Siddiq (1978). This material was used to examine the applicability of DNA markers for *badh2.1* and *badh2.7* for detecting fragrance allele types.

Identification of fragrant alleles by DNA markers

Genomic DNA was extracted from 2- week-old rice seedlings using 3 cm long leaf samples according to the method described by Gimhani *et al*. (2012). The leaf samples were homogenized with 300 µl of DNA extraction buffer (1 M KCl, 1 M Tris HCl, 0.5 M EDTA) and the homogenized samples were incubated at 70°C for 20 min. They were centrifuged at 13,000 rpm for 15 min at room temperature and the supernatant retained. DNA was precipitated by adding 100 µl of ice cold iso-propanol to the supernatant followed by incubation at 4°C for 15-30 min. The DNA was pelleted by centrifugation at 13,000 rpm for 15 min at room temperature. DNA pellets were washed with 150 µl of 70 % ice cold ethanol by centrifuging at 13,000 rpm for 10 min. The supernatants were removed and the air dried pellets dissolved in 200 µl of 1/10 TE buffer (10 mM Tris, 1 mM EDTA) were used for PCR.

For *badh2.1* (*fgr*) genotyping, PCR was conducted using four primers, namely, *ESP* (TTGTTTGGAGCTTGCTGATG), *IFAP* (CATAGGAGCAGCTGAAATATATACC), *INSP* (CTGGTAAAAAGATTATGGCTTCA) and *EAP* (AGTGCTTTACAAAGTCCCGC) by multiplexing in a single tube using the method of Bradbury *et al*. (2005b). Figure S1 (a) shows the primer annealing locations and amplification profile expected from *badh2.1* and wildtype alleles. For *Bad2.7CAPS* genotyping, PCR was performed in a total reaction volume of 15 µL, consisting

of 5× PCR buffer, 2.5 mM dNTPs, 25 mM $MgCl_2$, 0.2 µM *Bad2.7* primer (*Bad2.7CAPS*-Forward: 5'-CAAGTGAAGGGGATTG-3' and *Bad2.7CAPS*-Reverse: 5'-CCAAAGGCATGATGTCAGGTCG-3'), 5 u (1µL) Taq DNA polymerase and 20 ng of genomic DNA. Thermal cycle parameters included the following: an initial denaturation at 95 ^{0}C for 1 min, 35 cycles of 95 ^{0}C for 30 s, 56.9 ^{0}C for 30 s and 72 ^{0}C for 1 min, followed by a final extension at 72 ^{0}C for 7 min.

The amplified PCR products were digested by *Bsl*I restriction enzyme (BioLabs, China), which is a thermostable type II restriction endonuclease with interrupted recognition sequence CCNNNNN/NNGG (/, cleavage position) (Hsieh *et al*., 2000). Digestion reaction was carried out with 10 µL of PCR product in 16 µL of nuclease free water, 2 µL of 10 X buffer and 2 µL of 10 U/ µL *Bsl*I enzyme at 55 ^{0}C for 16 h, and the DNA was visualized in 3.0 % agarose gel. Figure S1 (b) shows the primer annealing locations and amplification profile expected from *badh2.7* and wildtype alleles.

Results

Prevalence of mutated alleles in exon regions of badh2 gene in Asian rice germplasm

This study searched for sequence variants in exons of the *badh2* gene across 1878 accessions (Table S1) in Rice SNP-Seek Database and identified 75 accessions with predicted loss of function mutations. The *badh2.1* allele was the most prevalent mutant allele occurring in 63 accessions while the *badh2.7* allele was the second most prevalent allele occurring in 9 accessions (Table 2

and Figure 2). In addition three novel non-functional alleles were observed in one accession each (Table 2).

Chung Yi (IRGC Acc. No. 1427-1) had several sequence mutations including G/C, C/G, T/A, C/G substitutions in exon 1, a C/T substitution in exon 2, A/G, A/T, T/A, T/G, T/G substitutions in exon 4, A/G, A/G, C/T substitutions in exon 5 and C/T, C/A, A/T, A/G, G/T, A/G, G/A, C/A substitutions in exon 10. The cDNA nucleotide sequence for this accession is predicted to generate a truncated 333 amino acid sequence resulting in an incomplete BADH2 protein. *Chugoku 68 Hen* (IRGC Acc. No.72514-1), a Japanese rice accession contained mutations such as A/W, C/Y, C/Y, A/R, A/R substitutions in exon 5 and C/Y, C/M, A/W, A/R, G/K, A/R, G/R substitutions in exon 10 (R = G or A substitution, Y= T or C substitution, K= G or T substitution, M= A or C substitution, S=G or C substitution, W= A or T substitution). It is predicted to produce a truncated 306 amino acid sequence instead of full length 503 amino acids. The accession, *E Pluak* (IRGC Acc. No. 64297-1) of Thailand origin had a T deletion in the 9th exon that is predicted to induce premature stop codon giving rise to a BADH2 protein with only 359 amino acids.

Twenty seven accessions contained synonymous mutations in *badh2* exons that were not predicted to cause loss of function of the wild type allele (Table S1). Several substitutions were observed in exon regions such as substitution of C/T in exon 02, C/T in exon 05, A/C in exon 06, A/T in exon 07, G/A and C/A in exon 10, C/T in exon 12, G/A in exon 13, G/A in exon 14 and C/T in exon 14 in 9 countries analyzed in this study (Table S1). These synonymous silent mutations were not considered of value to breeders.

The *badh2.1* allele was present in accessions from Bangladesh, Cambodia, India, Indonesia, Laos, Malaysia, Myanmarr, Nepal, Pakistan, Phillipines, Thailand and Vietnam while *badh2.7* allele was present only in accessions from India, Bangladesh, Pakistan and Sri Lanka. There are 47 accessions

with Sri Lankan origin in the rice SNP-Seek Database, out of which only *Kurulu Wee* (IRGC Acc. No. 66615-1) can be considered as a fragrant rice variety due to presence of the *badh2.7* allele. Accessions from countries such as Afghanistan, Butan, Iran, North Korea, South Korea, Taiwan and Tibet did not contain any fragrant rice accession that possessed the mutations reported in Table 1.

Variations of intron regions of the badh2 gene

A number of substitutions and *Indels* were detected in the intron regions of the *badh2* fragrance gene in the rice accessions analyzed in comparison to GenBank Acc. No. EU770319.1 (Data not shown). These are considered to have no effect on the expression of BADH2. We noted at least one mutation in at least one intron in each of 600 accessions tested in this study, however they were unlikely to affect the splicing process as they were present at the middle of the introns.

Identification of germplasm with *badh2.1* allele

This study used 28 rice lines (Table 3) including traditional varieties, improved varieties and advance breeding lines to compare fragrance phenotype with genotype using previously reported fragrance allele specific markers.

The *badh2.1 frg* marker originally developed by Bradbury *et al*. (2005b) amplified the 257 bp band for fragrant allele in varieties *MA2*, *At 309* and *At 311* containing *badh2.1* allele (Table 3; selected varieties shown in Figure 3A). None of Sri Lankan traditional rice contained the *badh2.1* allele although it is the most predominant allele in the South Asian region. The varieties, *MA2*, *At 309* and *At 311* are improved varieties which had been selected from crosses with a *Basmati* variety which could be the donor of *badh2.1* allele. As expected, non-fragrant rice varieties, *Bg 300* and

Bg 360 produced wild type alleles for the locus. Hence, the *badh2.1* (*frg*) marker comprising of *ESP, IFAP, INSP* and *EAP* primers is reliable for discrimination of *badh2.1* from its wild type, however it clearly is not a definitive test for the fragrant phenotype in all rice.

Validation of *Bad2.7CAPS* marker with fragrant rice germplasm

Dissanayaka *et al*. (2014) developd a CAPS marker that could discriminate *badh2.7* allele from its wild type. This is a two-step marker, requiring amplification of the target 457 bp DNA fragment followed by restriction digestion. The digest run on an agarose gel shows two possible patterns depending on the allele present (Figure 3B). Rice accessions that did not have G insertion in 14th exon region (wild type) produced 3 bands (255 bp, 67 bp and 121 bp) due to the presence of two *Bsl*I restriction sites. Rice accessions which had the *badh2.7* allele produced only two visible bands (188 bp and 255 bp) due to presence of the G insertion that disrupted *Bsl*I restriction site (In the wild type sequence, the *Bsl*I restriction site was present in the 188 bp fragment so it was split into two, 67 bp and 121 bp). Therefore, rice *Inginimitiya, Suwandel, and Suwanda samba* varieties possessing the fragrant *badh2.7* allele exhibited two DNA bands due to nucleotide G insertion that disrupted *Bsl*I site (Figure 3B). Advanced breeding lines of the *Suwandel* × *Bg 360* cross did not show either the mutant *badh2.7* allele or the fragrance phenotype while advanced lines of the *Iginimitiya* × *Bg 300* cross exhibited both the fragrance phenotype and the mutant *badh2.7* genotype (Table 3) indicating the applicability of *Bad2.7CAPS* marker for the selection of fragrant trait in these segregating populations. However, traditional rice variety, *Suduru samba* did not exhibit either *badh2.1* or *badh2.7* fragrant alleles in spite of its pleasant fragrance.

Discussion

The present study revealed that 72 rice accessions conserved at the International Rice Germplasm Collection contain either the common *badh2.1* allele or the rare *badh2.7* allele (Table 2). Kovach *et al*. (2009) and Dissanayaka *et al.* (2014) have shown that *badh2.1* and *badh2.7* mutant alleles critically affect the three dimensional structure of the BADH2 protein. Therefore, it is likely that aroma in aromatic rice varieties containing *badh2.1* and *badh2.7* is due to enhanced 2AP levels due to a non-functional BADH2 enzyme. Kovach *et al*. (2009) showed that, rice accessions possessing these two alleles produce 2AP, the major aroma causing compound. Therefore, the 72 accessions revealed in this study could be used in combination with the markers for *badh2.1* and *badh2.7* as parents in fragrant rice breeding.

Moreover, this study revealed a SNP mutation, T deletion in 09th exon, in the Thai rice variety, *E Pluak* (IRGC Acc. No. 64297-1), which has not been reported previously and therefore, it can be speculated that *E Pluak* could be a novel potential fragrant rice variety. Two other novel mutation types were detected in the Chinese variety *Chung Yi* (IRGC No. 1427-1) and in the Japanese variety *Chugoku 68 Hen* (IRGC Acc. No.72514-1). Further studies on sequence verification and phenotype assessment are necessary to confirm these three novel mutation types and the relationship between these alleles and fragrance.

Based on the variants detected in diverse genotypes it can be speculated that the *badh2.7* allele is limited only to rice varieties originating in the South Asian region while the *badh2.1* allele is found among varieties originating from countries of both Southeast Asia and South Asia. Although the *badh2.1* was the most prevalent allele in accessions of Asian origin, none of the 47 Sri Lankan

varieties possessed this mutation. This may be due to the fact that those varieties originated at the most southern part of Asia, where gene flow based transfer of *badh2.1* might not have occurred. The four countries (Pakistan, India, Bangladesh and Sri Lanka) containing rice germplasm with *badh2.7* allele are geographically close to each other, suggesting that one possible route for gene flow for the *badh2.7* allele was from a common ancestor that evolved in this region. Our results also support the theory that the *badh2.1* allele could have originated in japonica varietal group while *badh2.7* could have originated with in indica varietal group as hypotheised by Kovach *et al*. (2009), due to geographic distribution pattern of the alleles among countries of origin.

Mutations in intron regions such as a TT deletion in intron 2 and a repeated AT insert in intron 4 have been reported by (Chen *et al*., 2008) in Local Chinese fragrant rice varieties. However, whether the particular mutations in introns affected on fragrance, was not reported. Mukherjee *et al*. (2018) have observed high rate of accumulation of SNPs in introns as compared to exons in *Arabidopsis thaliana*. Moreover, they found a strong negative correlation between total exonic SNP density and intron number indicating that introns may protect the exonic regions against the harmful effects from such mutations. There are few reports illustrating the impact of mutations in intron regions in plants because usually introns do not contribute for protein structure, unless affecting the splicing process. Therefore, further studies are necessary to observe the impact of the variations in intron regions observed in this study on fragrance.

Breeding of fragrant rice consumes much time and needs skilled labor for the evaluation of fragrance. This task can be achieved more easily if a DNA marker that is tightly linked with fragrance is used in Marker Assisted Breeding (MAB) programs. There are many reports of successful use of fragrant allele specific markers in MAB (Shi *et al*., 2008; Hashemi *et al*., 2015;

He *et al*., 2015 and Peng *et al*., 2018). We tested local germplasm to detect the *badh2.1* fragrant allele using the *frg* (Bradbury *et al*., 2005b) DNA marker. The results confirm that this *badh2.1* marker is practically convenient to discriminate *badh2.1* from its wild type, but that they are not suitable for discrimination of other variants for aroma.

In Sri Lanka, At405, At306, At309, At311, At373 are improved high yielding fragrant varieties which inherited the pedigree of Basmati or IR 70422-66-5-2 (IRRI variety), both of which contain the *badh2.1* fragrant allele. More recently, the Rice Research and Development Institute, Sri Lanka focused on breeding programs with local traditional varieties, targeting pleasant fragrance and adaptability characters which could be obtained from Suwandel and Inginimiiya. However, due to lack of the availability of DNA markers, Suwandel and Inginimiiya could not be employed with expected rapidity via MAB programs. In the present study we employed the *Bad2.7CAPS* marker to detect the *badh2.7* variant in unexploited traditional aromatic varieties in order to inform the potential in MAB. Three traditional rice germplasm conserved at the genebank of RRDI, Sri Lanka (*Iginimitiya*, *Suwandel* and *Suwanda samba*) were found to have *badh2.7* allele at 14th exon (Figure 3 and Table 3) and screening was continued with progression of MAB. Also, it was observed that traditional rice variety, *Suduru samba* exhibited neither *badh2.1* nor *badh2.7* alleles so its fragrance may be due the presence of another unknown allele in its genome – either in *badh2* or another gene.

In breeding programs of rice, molecular markers are the preferred method to select for the fragrant trait at the early growth stage to reduce the time and cost of breeding. Although, there are reports of successful developments of aromatic rice by MAB using *badh2.1* markers (Shi *et al*., 2008 and Hashemi *et al*., 2015), no such breeding efforts have been reported for *badh2.7* based DNA

markers. As *badh2.7* is the second most prevalent allele in Asian region, this marker has potential to be used in the discovery of unutilized fragrant varieties and accelerate breeding if varieties carrying the allele are used as parents.

The majority of rice consumers prefer fragrant rice over non fragrant if other traits are matched. However, only a few fragrant rice varieties such as the *Basmati* varieties and Thai *Jasmine* varieties have become highly popularized and therefore, they are sold at a premium price in the market. If there are more varietal choices for fragrant rice, then farmers and breeders will gain the opportunity to cross them with other suitable parents to develop new fragrant varieties which could ultimately lead to the supply of more fragrant rice varieties at a competitive price. For sustainable rice production, it is necessary to breed varieties that are best suited to the environment of a particular country because when popular varieties such as *Basmati* are introduced to other regions they can be reported to have poorer growth and yield compared to their maximum yield in their native countries, Pakistan and India. Therefore, novel fragrant germplasm identified using the methods described in this study is likely to have as yet untapped value for breeding or cultivation by rice consuming countries.

Conclusion

In conclusion, this study revealed 75 fragrant accessions out of 1878 rice accessions originating from different Asian countries, available at International Rice Germplasm Collection, IRRI. The most prevalent fragrant allele in Asia was *badh2.1*, present in 63 accessions, while the second most prevalent allele was *badh2.7*, present in 9 accessions. Three accessions that possessed new allelic variations with the potential of causing fragrance have also been found, including one of which had a T deletion in the 9th exon. There is now potential for the development of molecular markers

for these novel variants. Presence of nucleotide variants at least in one intron of the *badh2* gene was a common feature of all tested accessions. The screening systems of *badh2.1* and *badh2.7* alleles were validated using Sri Lankan germplasm. The marker, *Bad2.7CAPS* was confirmed to be capable of discriminating fragrant and non-fragrant phenotypes using traditional and advanced breeding lines of rice.

Acknowledgement

Authors acknowledge NRC-09-11 research grant funded by National Research Council, Sri Lanka.

Figure 1: Schematic diagram showing the exon regions of *badh2* gene (www.ncbi.nlm.nih.gov/nuccore/eu770319.1) and their previously reported mutations as indicated in Table 1.

Figure 2: Multiple Sequence Alignment showing *badh2* variants for accessions originating from Sri Lanka and India in the rice SNP seek database; A: Accessions with a G insertion in the 14th exon indicated by * (*badh2.7*). B: Accessions with three SNPs and an 8 bp deletion in the 7th exon (*badh2.1*).

Figure 3: Molecular marker profiles for differentiation of fragrant alleles.

A: *badh2.1* allele using marker developed by Bradbury *et al*. (2005); B: *badh2.7* allele using *Bad2.7CAPS* marker; lane 1: *Iginimitiya* (aromatic); lane 2: *Suwandel* (aromatic); lane 3: *Suwanda samba* (aromatic); lane 4: *Suduru samba* (aromatic); lane 5: *Bg 300* (non-aromatic); lane 6: *At 311* (aromatic); lane 7: *At 309* (aromatic); lane L: 100 bp ladder.

Figure S 1: Schematic diagrams showing PCR profiles generated from *badh2.1* and *badh2.7* alleles

a) PCR profile generated from *badh2.1* allele using multiplexing with External sense primer (ESP), External antisense primer (EAP), Internal non-fragrant sense primer (INSP), Internal fragrant antisense primer (IFAP) (Bradbury *et al*.,2015b).

   The 355 bp band corresponds to a PCR product amplified from the non-fragrant allele by the INSP and EAP while the 257 bp band corresponds to a PCR product amplified from the fragrant allele by the IFAP primer ESP. The 580 or 577 band corresponds to the positive control amplified by ESP and EAP.

b) PCR profile generated from *badh2.7* allele using *Bad2.7CAPS* marker (Dissanayake *et al*., 2014)

   The 255 bp and 188 bp bands corresponds to the DNA fragment amplified by *Bad2.7CAPS* primers followed by *Bsl*I restriction digestion in the presence of 'G' base insertion (*badh2.7*allele). The 255 bp, 67bp and 121bp bands correspond to the DNA fragment amplified by *Bad2.7CAPS* primers followed by *Bsl*I restriction digestion in a variety with wild allele.

References

1. Alexandrov N, Tai S, Wang W, Mansueto L, Palis K, Fuentes RR and Mauleon R (2014) SNP-Seek database of SNPs derived from 3000 rice genomes. *Nucleic acids research* 43(D1): D1023-D1027.
2. Amarawathi Y, Singh R, Singh AK, Singh VP, Mohapatra T, Sharma TR and Singh NK (2008) Mapping of quantitative trait loci for *Basmati* quality traits in rice (*Oryza sativa* L.*). Molecular Breeding* 21(1): 49-65.
3. Bindusree G, Natarajan P, Kalva S and Madasamy P (2017) Whole genome sequencing of Oryza sativa L. cv. Seeragasamba identifies a new fragrance allele in rice. *PloS one* 12(11): e0188920.
4. Bradbury LM, Fitzgerald TL, Henry RJ, Jin Q and Waters DL (2005a) The gene for fragrance in rice. *Plant Biotechnology Journal* 3(3): 363-370.
5. Bradbury LMT, Henry RJ, Jin Q, Reinke RF and Waters DLE (2005b) A perfect marker for fragrance genotyping in rice. *Molecular Breeding* 16: 279-283.
6. Buttery RG, Ling LC and Juliano BO (1982) 2-Acetyl-1pyrroline: An important aroma component of cooked rice. *Chemistry and Industry (London)* 23: 958-959.
7. Buttery RG, Ling LC, Juliano BO and Turnbaugh JG (1983) Cooked rice aroma and 2-acetyl-1-pyrroline. *Journal of Agricultural and Food Chemistry* 31(4): 823-826.
8. Bourgis F, Guyot R, Gherbi H, Tailliez E, Amabile I, Salse J and Ghesquiere A (2008) Characterization of the major fragrance gene from an aromatic japonica rice and analysis of its diversity in Asian cultivated rice. *Theoretical and Applied Genetics* 117(3): 353-368.
9. Chen S, Yang Y, Shi W, Ji Q, He F, Zhang Z and Xu M (2008) *Badh2*, encoding betaine aldehyde dehydrogenase, inhibits the biosynthesis of 2-acetyl-1-pyrroline, a major component in rice fragrance. *The Plant Cell* 20(7): 1850-1861.

10. Cordeiro GM, Christopher MJ, Henry RJ and Reinke RF (2002) Identification of microsatellite markers for fragrance in rice by analysis of the rice genome sequence. *Molecular Breeding* 9: 245-250.
11. Dissanayaka S, Kottearachchi NS, Weerasena J and Peiris M (2014) Development of a CAPS marker for the *badh2. 7* allele in Sri Lankan fragrant rice (*Oryza sativa*). *Plant Breeding* 133(5): 560-565.
12. Gimhani DR, Kottearachchi NS, and Samarasinghe WLG (2014). Microsatellite marker based hybridity assessment; An approach towards development of mapping population for salinity tolerance in rice. Journal of agricultural sciences. (9)2, 96 – 100.
13. Giraud G (2013) The world market of fragrant rice, main issues and perspectives. *International Food and Agribusiness Management Review* 16(2): 1-2.
14. Hashemi FSG, Rafii MY, Ismail MR, Mohamed MTM, Rahim HA, Latif MA and Aslani A (2015) Opportunities of marker-assisted selection for rice fragrance through marker-trait association analysis of microsatellites and gene-based markers. *Plant Biology* 17(5): 953-956.
15. He Q and Park YJ (2015) Discovery of a novel fragrant allele and development of functional markers for fragrance in rice. *Molecular Breeding* 35(11): 217.
16. Hinge V R, Patil HB and Nadaf AB (2016) Aroma volatile analyses and 2AP characterization at various developmental stages in *Basmati* and Non-*Basmati* scented rice (*Oryza sativa* L.) cultivars. *Rice* 9(1): 38.
17. Hsieh P, Xiao J, Oloane D and Xu S (2000) Cloning, Expression and Purification of a Thermostable Nonhomodimeric Restriction Enzyme, *Bsl*I. *Journal of Bacteriology*, 182 (4) 949–955.

18. Kovach MJ, Calingacion MN, Fitzgerald MA and McCouch SR (2009) The origin and evolution of fragrance in rice (*Oryza sativa* L.). *Proceedings of the National Academy of Sciences* 106(34): 14444-14449.
19. Kumar S, Stecher G and Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular biology and evolution* 33 (7): 1870-1874.
20. Lorieux M, Petrov M, Huang N, Guiderdoni E and Ghesquiere A (1996) Aroma in rice–genetic analysis of a quantitative trait. *Theoretical and Applied Genetics* 93: 1145-1151.
21. Mukherjee D, Saha D, Acharya D, Mukherjee A, Chakrabort S and Ghosh TC (2018) The role of introns in the conservation of the metabolic genes of *Arabidopsis thaliana*. *Genomics* 110(5): 310-317.
22. Peng B, Zuo YH , Hao YL , Peng J , Kong DY, Peng Y, Nassirou TY, He LL , Sun YF, Liu L, Pang RH, Chen YX, Li JT, Zhou QY, Duan B, Song XH , Song SZ and Yuan HY (2018) Studies on aroma gene and its application in rice genetics and breeding. *Journal of Plant Studies* 7(2): 29-40.
23. Pinson SRM (1994) Inheritance of aroma in six rice cultivators *Crop Science* 34: 1151–1157.
24. Sakthivel K, Sundaram RM, Rani NS, Balachandran SM and Neeraja CN (2009) Genetic and molecular basis of fragrance in rice. *Biotechnology Advances* 27(4): 468-473.
25. Shao G, Tang A, Tang S, Luo J, Jiao G and Wu J (2011) A new deletion mutation of fragrant gene and the development of three molecular markers for fragrance in rice. *Plant Breeding* 130(2): 172-176.
26. Shao G, Tang S, Chen M, Wei X, He J and Luo J (2013) Haplotype variation at *Badh2*, the gene determining fragrance in rice. *Genomics* 101(2): 157-162.

27. Shi W, Yang Y, Chen S and Xu M (2008) Discovery of a new fragrance allele and the development of functional markers for the breeding of fragrant rice varieties. *Molecular Breeding* 22(2): 185-192.
28. Sood B C and Siddiq EA (1978) A rapid technique for scent determination in rice. *Indian Journal of Genetics and Plant Breeding* 38(2): 26.

Table 1: Details of the mutations in exons regions of *badh2* gene analyzed in this study

Exon Region	Allele Name	Mutations	References
Exon 1	*badh2.3*	2 bp deletion	Kovach *et al*. (2009)
Exon 2	*badh2.2*	7 bp deletion	Kovach *et al*. (2009)
	(*badh2-E2*)		Shi *et al*. (2008)
	badh2-E2.2	75 bp deletion	Shao *et al*. (2013)
Exon 4 to Exon 5	*badh2-E4-5.1*	806 bp deletion	Shao *et al*. (2013)
	badh2-E4-5.2	803 bp deletion	Shao *et al*. (2011)
Exon 7	*badh2.1*	8 bp deletion with 3 SNPs	Bradbury *et al*. (2005)
Exon 8	*badh2-E8*	7 bp insertion	Amarawathi *et al*. (2008)
Exon 10	*badh2.5*	1 bp deletion (T)	Kovach *et al*. (2009)
	badh2.6	substitution (G → T)	Kovach *et al*. (2009)
	badh2.4	1 bp insertion	Kovach *et al*. (2009)
	badh2-E10.4	G/T SNP	Shao *et al*. (2013)
Exon 12	*badh2-p*	3 bp deletion	He *et al*. (2015)
Exon 13	*badh2.8*	1 bp deletion (T)	Kovach *et al*. (2009)
	badh2-E13.1	3 bp insertion	Kovach *et al*. (2009)
	badh2.10	substitution (C → T)	Kovach *et al*. (2009)
Exon 14	*badh2.7*	1 bp insetion (G)	Kovach *et al*. (2009)
	badh2.9	substitution (G → T)	Kovach *et al*. (2009)
	*	substitution (C→ T)	Bourgis *et al*. (2008)

* Name has not been given for the allele

Table 2: Rice accessions detected with mutations in *badh2* gene

IRGC No	Variety Name	Allele	IRGC No	Variety Name	Allele
Bangladesh			107773-1	MEE	*badh 2.1*
37793-2	BADSHABHOG 4-60	*badh 2.1*	**Malaysia**		
29260-1	BEGUNBICHI 33	*badh 2.1*	60310-1	KENDINGA 5 (H)	*badh 2.1*
29333-1	GASMAL 339	*badh 2.1*	28694-2	ENGKABANG	*badh 2.1*
83865-1	KASHA	*badh 2.1*	71514-1	DARAWAL	*badh 2.1*
29377-1	PANKAIT 31	*badh 2.1*	60392-1	PAMPANGON (H)	*badh 2.1*
53541-1	LALDIGHA	*badh 2.1*	71556-1	LIMBAYAN	*badh 2.1*
29230-1	AUS 449	*badh 2.7*	**Myanmar**		
Cambodia			70763-1	PAKISTANHMWE	*badh 2.1*
22841-2	DAMNOEUB POR HING	*badh 2.1*	70746-2	KHAOSAING	*badh 2.1*
22819-2	DAMNOEUB KHSE SAUT	*badh 2.1*	**Nepal**		
China			62043-1	SETO JHINUWA	*badh 2.1*
1427-1	CHUNG YI	[a]	16138-2	MADHUWA KARIA	*badh 2.1*
India			59205-1	RATO BASMATI	*badh 2.1*
20570-2	ARC 7263	*badh 2.1*	62025-1	POHHERLIMASION	*badh 2.1*
42976-2	ARC 14709	*badh 2.1*	**Pakistan**		
45024-2	BAJAL	*badh 2.1*	27797-1	BASMATI SURKH 161	*badh 2.1*
45397-1	CR 44-1	*badh 2.1*	27798-1	BASMATI 1	*badh 2.1*
	JC 1	*badh 2.1*	27779-1	BARA PASHAWARI	*badh 2.1*
	JC 157	*badh 2.1*	27976-1	JHONA 101	*badh 2.7*
46117-2	KEYA NUNIA	*badh 2.1*	**Philippines**		
70840-1	HIRA NAKHI	*badh 2.1*	87131-1	GINAYANGGANG	*badh 2.1*
49850-2	LAWANGAI	*badh 2.1*	52887-1	CARAWI	*badh 2.1*
42538-1	ARC 7425	*badh 2.7*	44637-2	MINANOK	*badh 2.1*
21614-1	ARC 11751	*badh 2.7*	96124-1	MALAGKIT (ITIM)	*badh 2.1*
41811-1	ARC 14901	*badh 2.7*	53146-1	KINTOMAN (TOD DOY)	*badh 2.1*
61525-1	UPRH 58:	*badh 2.7*	57599-1	J 6 IR 438 (WC 694)	*badh 2.1*
52343-1	LOCAL BHAT	*badh 2.7*	79058-1	TUI (KIDAYAN)	*badh 2.1*
39709-1	RATNAGIRI 45-2	*badh 2.7*	26942-1	PANGETAN	*badh 2.1*
Indonesia			67444-1	MALAGKIT (PINELIPE)	*badh 2.1*
17925-2	KETAN NARIS	*badh 2.1*	8059-1	PARAY QIKUG	*badh 2.1*
18789-1	SERUNEN	*badh 2.1*	23342-1	T Q QI QINYUHAYUHA	*badh 2.1*
35446-2	RUJAK DANAN	*badh 2.1*	**Sri Lanka**		
25821-2	TARING PELUNDUK	*badh 2.1*	66518-1	KURULU WEE (WHITE)	*badh 2.7*
25812-1	SUNGKAI	*badh 2.1*	**Thailand**		
16650-1	TJERE SUGI	*badh 2.1*	55265-1	PRADOO DAENG	*badh 2.1*
66645-1	SIGAE	*badh 2.1*	48007-1	HAWM TOONG	*badh 2.1*
66630-1	SAHULO FACHE SOYO	*badh 2.1*	47938-1	E DAW HAWM	*badh 2.1*
Japan			66704-1	DAW 85	*badh 2.1*
72514-1	CHUGOKU 68 HEN	[b]	64297-1	E PLUAK	[c]
Laos			61359-1	YAH YAW	*badh 2.1*
84920-1	CHA LIENG	*badh 2.1*	71220-1	KHAO NAHNG PRUNG	*badh 2.1*
78776-1	CHON	*badh 2.1*	71001-1	KHAO' DAENG HAWM	*badh 2.1*
107021-1	DOK HIEN NOI	*badh 2.1*	**Vietnam**		
107709-1	MAK KHEUA DENG	*badh 2.1*	7082-2	LUA HUONG T 1	*badh 2.1*

[a]Several substitutions in exon 1, 2, 4, 5 and 10, [b] several substitutions in exon 5 and 10, [c] T deletion in exon 9

Table 3: Types of *badh2* alleles detected using allele specific markers

Variety or breeding line	Allele at 7^{h} exon	Allele at 14th exon
Iginimitiya (traditional)*	wild	*badh 2.7*
Suwandel (traditional)*	wild	*badh 2.7*
Suduru samba (traditional)*	wild	wild
Suwanda samba (traditional)*	wild	*badh 2.7*
MA2 (improved)*	*badh2.1*	wild
At 309 (improved)*	*badh2.1*	wild
At 311(improved)*	*badh2.1*	wild
Bg 300 (improved)	wild	wild
Bg 360 (improved)	wild	wild
P360-L1[a]*	wild	*badh2.7*
P360-L2[a]*	wild	*badh2.7*
P361-L1[a]*	wild	*badh2.7*
P361-L2[a]*	wild	*badh2.7*
P362-L1[a]*	wild	*badh2.7*
P362-L2[a]*	wild	*badh2.7*
P363-L1[a]*	wild	*badh2.7*
P363-L2[a]*	wild	*badh2.7*
P363-L3[a]*	wild	*badh2.7*
P363-L4[a]*	wild	*badh2.7*
P382-L2[b]	wild	wild
P383-L2[b]	wild	wild
P385-L2[b]	wild	wild
P386-L1[b]	wild	wild
P386-L2[b]	wild	wild
P386-L3[b]	wild	wild
P386-L4[b]	wild	wild
P402-L1[b]	wild	wild
P402-L3[b]	wild	wild

*Aromatic lines; [a] Advance breeding progeny lines of *Iginimitiya* × *Bg 300*; [b] Advance breeding progeny lines of *Suwandel* × *Bg 360*